UC Davis

UC Davis Previously Published Works

Title

The Affect Misattribution Procedure.

Permalink

https://escholarship.org/uc/item/5n83n22p

Journal

Experimental psychology, 64(3)

ISSN

1618-3169

Authors

Teige-Mocigemba, Sarah
Becker, Manuel
Sherman, Jeffrey W
et al.

Publication Date

2017-05-01

DOI

10.1027/1618-3169/a000364

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Registered Report

The Affect Misattribution Procedure

In Search of Prejudice Effects

Sarah Teige-Mocigemba,[1] Manuel Becker,[1] Jeffrey W. Sherman,[2] Regina Reichardt,[3] and Karl Christoph Klauer[1]

[1]Institut für Psychologie, Sozialpsychologie und Methodenlehre, Albert-Ludwigs-Universität Freiburg, Germany
[2]Department of Psychology, University of California, Davis, CA, USA
[3]Institut für Psychologie, Lehrstuhl für Pädagogische Psychologie, University of Regensburg, Germany

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Abstract: The Affect Misattribution Procedure (AMP) has been forwarded as one of the most promising alternatives to the Implicit Association Test and the evaluative-priming task for measuring attitudes such as prejudice indirectly. We investigated whether the AMP is indeed able to detect an evaluative out-group bias. In contrast to recent conclusions about the robustness of AMP effects, six out of seven pilot studies indicated that participants did not show any prejudice effects in the AMP. Yet, these pilot studies were not fully conclusive with regard to our research question because they investigated different domains of prejudice, used small sample sizes, and employed a modified AMP version. In a preregistered, high-powered AMP study, we therefore examined whether the standard AMP does reveal prejudice against Turks, the biggest minority in Germany, and found a significant, albeit very small prejudice effect. We discuss possible reasons for the AMP's weak sensitivity to evaluations in socially sensitive domains.

Keywords: affect misattribution procedure, implicit measures, prejudice effects

In the last two decades, a number of so-called "indirect" or "implicit" measures have been developed to measure attitudes unobtrusively (for a recent overview, see, e.g., Gawronski & De Houwer, 2014). In contrast to self-report measures, indirect measures do not directly ask participants to provide information about themselves. Instead, this information is inferred from participants' response latencies, error rates, or response frequencies because task performance is assumed to be shaped by the to-be-measured construct (e.g., an attitude; De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009). The indirect measurement approach is hoped to circumvent the two key problems of self-report measures, namely introspective limits (cf. Nisbett & Wilson, 1977) and distortions by self-presentation tendencies (e.g., Fazio, Jackson, Dunton, & Williams, 1995).

Although past research has revealed that indirect measures only rarely possess all features of automaticity that are typically ascribed to them (see De Houwer et al., 2009), their utility has been shown in a multitude of studies (for a comprehensive overview, see the Handbook of Implicit Social Cognition edited by Gawronski & Payne, 2010). For instance, the incremental value of indirect measures has been demonstrated in studies showing that indirect measures and self-report measures respond differently to the same experimental manipulations (e.g., Gawronski & Strack, 2004) and predict different aspects of behavior (e.g., Asendorpf, Banse, & Mücke, 2002; Cameron, Brown-Iannuzzi, & Payne, 2012; Fazio et al., 1995; Greenwald, Banaji, & Nosek, 2015; Greenwald, Poehlman, Uhlmann, & Banaji, 2009; but see Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013, 2015), while being only weakly related to each other under many conditions (e.g., Hofmann, Gschwendner, Nosek, & Schmitt, 2005). In a similar vein, at the group level, indirect measures have been shown to reveal a priori expected preferences (e.g., a preference for the in-group) that are sometimes concealed in self-report measures (e.g., Fazio et al., 1995; Jones & Sigall, 1971). In fact, research on indirect measures has often used the latter approach to attest to an indirect measure's validity (see De Houwer et al., 2009).

Among the currently most frequently used indirect measures rank the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), the evaluative-priming task (EPT; Fazio et al., 1995), and the Affect Misattribution Procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005). Both the IAT and the EPT have been subject to some criticism. For example, due to its task structure, the IAT has been argued to be susceptible to various sources of (un)systematic construct-unrelated variance, although

© 2017 Hogrefe Publishing

DOI: 10.1027/1618-3169/a000364

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

it typically shows satisfactory internal consistencies and large effect sizes (Teige-Mocigemba & Klauer, 2015; Teige-Mocigemba, Klauer, & Sherman, 2010). The EPT, on the other hand, has been criticized for its usually low internal consistencies and rather small effect sizes, whereas the EPT's task structure has been argued to prevent the influence of many variables known to confound IAT effects (see Wentura & Degner, 2010). In light of this criticism, the AMP has been forwarded as a promising alternative to the IAT and the EPT because it may combine the advantages of the IAT and the EPT (Gawronski & Ye, 2015). Specifically, the AMP has been shown to have satisfactory internal consistencies and large effect sizes, and it is structurally akin to sequential priming tasks that have a long tradition in social cognition research. Reviewing ten years of research on the AMP, Payne and Lundberg (2014) concluded that the AMP has become a well-validated tool.

Given these encouraging claims, we started to further investigate and use the AMP in our own research. In this paper, we report the results of these efforts. In six out of seven studies, we observed no evidence of prejudice (i.e., negative attitudes toward social groups) on the AMP. However, in a sample representative of our population of participants, IAT data and evaluative-priming data revealed the expected prejudice effects, suggesting that the AMP failed to detect prejudice when it was prevalent. A number of alternative explanations to this conclusion could be ruled out (see below). Given that the AMP is currently argued to be a particularly promising measure that produces robust effects, the present null findings are noteworthy and in need of empirical scrutiny (cf. Bakker, van Dijk, & Wicherts, 2012; Nosek & Lakens, 2014; Open Science Collaboration, 2015). We therefore conducted a final, preregistered, and high-powered AMP study to evaluate more conclusively the extent to which the AMP is sensitive to prejudice.

In what follows, we will first describe the basic AMP procedure, followed by a review of findings on the robustness of AMP effects. Subsequently, we will briefly report the general procedural details and the overall results of the seven pilot studies in which we searched for prejudice effects in the AMP. We will then provide the details of the preregistered, high-powered AMP study we conducted.

The AMP

In a standard AMP, participants are briefly presented with a prime stimulus (e.g., a positive or a negative image). Then, a blank screen appears and is replaced by a neutral target stimulus, typically a Chinese character, that is masked after 100 ms. Participants are asked to evaluate whether the Chinese character is visually more pleasant or more unpleasant than the average Chinese character. The AMP rests on the assumption that participants misattribute their prime evaluation to the evaluation of the target's pleasantness (e.g., Gawronski & Ye, 2014; Payne & Lundberg, 2014). Specifically, participants are expected to base their judgment of the target stimulus on their momentary mental states (e.g., negative affective feelings) that are assumed to be unobtrusively influenced by the preceding prime stimulus. For example, Chinese characters following negative primes (e.g., the picture of an aggressive dog) should be evaluated as less pleasant than Chinese characters following positive primes (e.g., the picture of a cute kitten) because compared to positive primes, negative primes activate more unpleasant affective feelings that should then influence target evaluations.

The AMP has been shown to produce robust effects across a wide range of attitudes (e.g., Gawronski, Cunningham, LeBel, & Deutsch, 2010; Payne et al., 2005; Payne & Lundberg, 2014). For instance, AMP effects occurred despite instructions to not let the primes influence responses to the targets (Payne et al., 2005). In a similar vein, AMP effects were found even when participants reported in post-experimental interviews that they had corrected for the influence of the primes (Payne et al., 2005). Finally, AMP effects were not influenced by social pressure manipulations that should motivate participants to hide their (alcohol-related) preferences (e.g., Payne, Govorun, & Arbuckle, 2008). On the other hand, AMP effects were shown to be larger for participants who later reported that the primes influenced their responses to the targets and that they had intentionally rated the primes instead of the targets (Bar-Anan & Nosek, 2012; but see Payne et al., 2013). Relatedly, AMP effects have been found to be weaker if participants were motivated to provide accurate pleasantness responses to the targets (Eder & Deutsch, 2015). Furthermore, if participants were requested to fake reversed attitudes in the AMP, they succeeded even without strategic advice and prior task experience (Teige-Mocigemba, Penzl, Becker, Henn, & Klauer, 2016). To summarize, past research suggests that in most cases, AMP effects are robust and large. The AMP can thus be expected to detect real differences in evaluations reliably. Under certain circumstances, however, AMP effects may be susceptible to strategic influences.

Seven Pilot Studies in Search of Prejudice Effects

Encouraged by the evidence of the utility of the AMP and its robust effects (e.g., Cameron et al., 2012; Payne & Lundberg, 2014), we aimed at assessing prejudice against minorities with the AMP. Particularly in these socially

 © 2017 Hogrefe Publishing

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

sensitive domains, indirect measures such as the AMP have been argued to reveal their utility (Greenwald et al., 2009). In this regard, we were originally interested in extending a recent multinomial process model of the AMP to estimate processes at the individual level instead of the group level only (Payne, Hall, Cameron, & Bishara, 2010). Starting from Payne and colleagues' (2010) model, we hypothesized that an individual-level version of the model would permit us to disentangle individual process estimates for misattribution (parameter M which reflects the extent of confusing prime and target evaluations), attitudes (parameter A which reflects the affective responses to the primes provided that misattribution occurred), and target evaluations (parameter P which reflects the affective responses to the targets provided that misattribution did not occur), respectively. Accordingly, we expected the individual process parameter A to reflect a less confounded estimate of attitudes (here, prejudice against minorities) that could be used for subsequent correlational analyses.

In what follows, we report the general method and the overall results of all seven AMP studies. In the Electronic Supplementary Material (ESM 1), more detailed information about the participants, the method, and the results of each study can be found.

Method

Procedural Details of the AMP

In order to apply Payne et al.'s (2010) process model to AMP data, some procedural deviations from the standard AMP are necessary. Most importantly, the duration of the target presentation has to be manipulated (in two steps: 100 ms vs. 1,000 ms) to have enough degrees of freedom for estimating and testing the model. Regarding the procedural details of the AMP, we followed Payne et al. (2010) as closely as possible, with one important exception: We manipulated the duration of target presentation within participants, not between participants, because we aimed at estimating an individual-level model, not a group-level model.

Primes

Prime stimuli were portraits representing three prime categories, respectively: a neutral prime category, an out-group prime category, and an in-group prime category. We included a neutral prime category because it is common to do so in race AMPs (see Payne & Lundberg, 2014) and because we wanted to be able to reanalyze the data with an alternative multinomial process model developed for a stereotype variant of the AMP (Krieglmeyer & Sherman, 2012) that requires neutral prime trials. In all AMP studies, four gray, face-like shapes provided by Krieglmeyer and Sherman (2012) served as primes for the neutral prime category. As out-group prime categories, we selected different groups known to face prejudice in Germany. Specifically, the out-group and in-group prime categories were Turks and Germans (Study 1), Blacks and Whites (Studies 2 and 6), old and young women (Study 3), Caucasian women wearing traditionally tied Muslim headscarves and the same women bareheaded (Studies 4 and 7), and Arabs associated with terrorism and liked German celebrities (Study 5). Out-group and in-group prime categories were represented by four portraits each.

Following Payne et al. (2010; see Footnote 2), great efforts were made to equate the stimulus sets of the three prime categories per study, with the exception of Study 5 that used non-equated stimulus sets from prior research (see Experiment 1 of Teige-Mocigemba & Klauer, 2013). Out-group and in-group prime stimuli were selected such that the portrayed persons of the two stimulus sets were (a) unequivocally identifiable as representatives of their respective groups and (b) matched for attractiveness and valence of emotional expression (for details, see the ESM). This was done to ensure that the expected AMP effects (i.e., more positive evaluations of targets following in-group primes than of targets following out-group primes) indeed result from differences in group evaluations, and not from differences in attractiveness or emotional expression that may contribute to or override effects of prejudice. For instance, if the portrayed persons of the in-group prime category would be more attractive than those of the out-group prime category, AMP effects may be completely accounted for by these differences in attractiveness (see Footnote 2 of Payne et al., 2010). In a similar vein, Craig, Lipp, and Mallan (2014) recently showed that differences in emotional expressions elicited evaluations that were strong enough to override the a priori expected effects of prejudice against Blacks and against old people on an EPT.

Stimulus selections relied on available pre-ratings (Studies 1 and 3) or our own pre-rating studies (Studies 1, 2, and 6), whereas Studies 4 and 7 used materials that inherently prevented any confounding influence by using the same persons as exemplars of both categories, once equipped with headscarves characteristic of Muslim women, once not equipped with headscarves (non-Muslim women). Note that, for each study, the neutral primes were edited in terms of background color and shirt (as well as shirt color) to make them as similar to the out-group and in-group prime pictures as possible.

Targets

In order to estimate Payne et al.'s (2010) process model, targets have to be selected that are ambiguous enough to allow for the misattribution process while being valenced enough to be classified as (slightly) more pleasant or (slightly) more unpleasant than the average Chinese character. Like Payne and colleagues (2010; see Footnote 1),

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

we based our target selections on reanalyses of past AMP data. Note that we did not use Payne et al.'s target selections because we conducted our studies in Germany and, therefore, wanted to select targets according to their evaluations in a German sample.

Following Payne et al.'s procedure, we reanalyzed the average target evaluations (independent of prime type) of the 77 participants of the control groups in Experiments 1 (n = 27) and 2 (n = 50) of Teige-Mocigemba et al. (2016). The target set comprised 192 different Chinese characters that we downloaded from Keith Payne's website (http://www.unc.edu/~bkpayne/materials.html) and that are typically used in a standard AMP. The rate of pleasant responses to these Chinese characters (averaged across participants) ranged from 29% to 86% (M = 54%, SD = 12%; descriptive statistics of Payne et al. were M = 57%, SD = 7%; range: 46%–75%).

For Study 1, we selected the 72 most negative Chinese characters (M = 45% pleasant responses) and the 72 most positive Chinese characters (M = 63% pleasant responses), overall M = 54%, SD = 10%. Due to a very strong main effect of target valence in Study 1 (see the ESM), we selected less valenced target stimuli for Studies 2–7. Specifically, we selected the 72 least negative Chinese characters from the lower half of the distribution of pleasantness ratings, resulting in a set with an average of 47% pleasant responses, and we selected the 72 least positive Chinese characters from the upper half of the distribution of pleasantness ratings, resulting in a set with an average of 61% pleasant responses. Across the two stimulus sets of slightly negative and slightly positive targets, there were M = 54% pleasant responses (SD = 8%), ranging from 42% to 70%. Target stimuli had a size of 256 × 256 pixels and all stimuli were shown on a white background on a 58 cm LCD TFT monitor (LG Flatron W2363D) with a refresh rate of 120 Hz.

AMP Trials

Following Payne et al. (2010), an AMP trial started with a prime picture presented for 75 ms in the center of the screen, followed by a blank screen presented for 125 ms. Then, a target appeared for 100 ms or 1,000 ms (the duration of target presentation was manipulated within participants in a block-wise manner, see below). If responses were slower than the duration of target presentation, the target was masked by a pattern mask consisting of black and white noise that remained on screen until participants responded. Responses were given by pressing the [Q] for visually more unpleasant responses and the [P] for visually more pleasant responses on a standard German QWERTZ keyboard. The labels "unpleasant" and "pleasant" were presented in the corresponding corners of the screen throughout the experiment.

Sequence of AMP Trials

Participants first completed a practice block, followed by six test blocks. The practice block consisted of six trials with one stimulus of each prime category (neutral, in-group, out-group) being paired with a slightly pleasant target stimulus and another stimulus of each prime category being paired with a slightly unpleasant target stimulus. With these restrictions, primes and targets were paired randomly. All stimuli of the practice block (two stimuli per prime category, six target stimuli) were additional stimuli that were presented only in the practice block.

The six test blocks comprised 24 trials each, with each of the 12 prime pictures being presented twice per block and each of the 144 target pictures presented only once throughout all test blocks. The AMP was programmed such that, per test block, each prime stimulus was once paired with a negative target and once paired with a positive target. Prime-target pairs were presented in randomized order. Because, originally, we planned to conduct hierarchical multinomial modeling analyses, we ensured that, across participants, all target stimuli were paired at least once with all prime stimuli, either in a block presenting target stimuli for 100 ms or in a block presenting target stimuli for 1,000 ms (conditions referred to as $\text{target}_{\text{short}}$ vs. $\text{target}_{\text{long}}$ in the following). These requirements of prime-target pairings made it necessary to collect data of at least 27 participants per study. Importantly, from test block to test block, the duration of target presentation changed within participants, and between participants, it was counterbalanced whether participants first worked through a test block with a short or with a long duration of target presentation (referred to as first $\text{block}_{\text{short}}$ vs. first $\text{block}_{\text{long}}$ in the following). Note that the duration of target presentation in the practice block always corresponded to that of the first test block.

Participants

For the reasons elaborated on above, we collected data of 27 participants per study with the exception of Study 3, in which only 25 subjects participated. Participants were mostly University of Freiburg students with different majors. They received partial course credit or €1.50 for participation. Following recommendations by Payne and Lundberg (2014), participants were excluded from data analyses, if (a) their use of one response key was a "far-out" value (Tukey, 1977) in the distribution of the total sample (i.e., if they pressed the right response key conspicuously often or rarely; overall one participant)[1] or if (b) participants

[1] According to Tukey's (1977) outlier criteria, "far-out" values in the distribution of a given variable (here, mean percentage of right key presses) are values that are below (above) the first (third) quartile minus (plus) three times the total sample's interquartile range of this variable.

 © 2017 Hogrefe Publishing

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

indicated that they were able to read Chinese characters (overall four participants). Based on these exclusion criteria, 182 participants of the overall sample of 187 participants entered the final analyses (see the ESM for details).

Procedure

Upon arrival at the laboratory, participants first worked through the AMP. In all studies except Studies 1 and 5, participants subsequently rated the prime stimuli on a five-point Likert scale with regard to emotional expression (1 = *very negative*, 5 = *very positive*) and attractiveness (1 = *very unattractive*, 5 = *very attractive*). Then, they were asked to report personal data (age, gender, handedness, major, familiarity with Chinese language, and whether their relatives or close friends belonged to the out-group under investigation;[2] only in Study 3 the latter question was not asked). Finally, participants were thanked, compensated for participation, and debriefed.

Overview of Results

As detailed in the ESM and in Table 1, we failed to find prejudice effects (i.e., more negative evaluations of targets following out-group primes than in-group primes) in six out of seven studies. In Studies 1–4, we investigated different social groups our student sample is typically prejudiced against. After having found no preference for Germans over Turks in Study 1, $t(25) = .87$, $p = .39$, Cohen's $d = 0.17$, we initially thought that participants may just not be prejudiced against the specific out-group of Turks. However, the AMP did not reveal a preference for Whites over Blacks (Study 2) either, nor for young women over old women (Study 3), nor for non-Muslim women over Muslim women (Study 4), all $|t| \leq 1.09$, $p \geq .29$, $d \leq 0.22$. Because Turks, Blacks, old people, and Muslim women wearing headscarves have been found to face prejudice in Germany (see below), we suspected that the changes in the AMP procedure necessary to conduct the individual process-model analyses (i.e., within-participants manipulation of duration of target presentation) may have altered the procedure to such a great extent that the AMP may not be sensitive to prime evaluations anymore.

We tested this possibility in Study 5, using prime stimuli that have been shown to produce prejudice effects with an EPT (e.g., Teige-Mocigemba & Klauer, 2013). Specifically, we contrasted two prime categories that are of opposing valence, but are not natural counterparts, by presenting rather grim looking Arabs associated with terrorism (with beards and turbans) and rather friendly looking liked German celebrities as prime stimuli. Validating the present AMP procedure, we found the expected AMP effect with more negative evaluations of the terrorist-associated Arabs than of the liked celebrities, $t(26) = 3.11$, $p = .004$, Cohen's $d = 0.60$. Based on these findings, we proceeded with the AMP procedure we had used in Studies 1–5.

In Study 6, we again investigated participants' prejudice against Blacks. Participants' post-experimental ratings of the out-group and in-group primes used in Study 2 had indicated that the portraits of Blacks were more attractive than those of Whites. Accordingly, attractiveness might have overridden prejudice effects, thereby explaining the null finding. We therefore selected new sets of prime stimuli that were better matched with regard to attractiveness and valence of emotional expression (for details, see the ESM). However, even with the newly selected prime sets, the AMP did not reveal the expected preference for Whites over Blacks, $t(26) = -.46$, $p = .65$, Cohen's $d = -0.09$. This is remarkable because in post-experimental ratings of the prime stimuli, the portraits of Whites were now even rated as more attractive and as showing a more positive emotional expression than the portraits of Blacks, which should have contributed to the expected prejudice effect in the AMP.

In a final attempt to find prejudice effects in the AMP, we hypothesized that for prejudice effects to occur, the relevant social categories must be salient (for a similar argument, see Craig et al., 2014; Olson & Fazio, 2003; but see Gawronski et al., 2010). As has been found for the EPT, the AMP, as well, might reflect spontaneous evaluations of the exemplars (e.g., the specific portraits) rather than of the social categories (e.g., the social group the portrayed person belongs to). In Study 7, we therefore increased category salience by means of a categorization task prior to the AMP (for a similar manipulation of category salience, see Deffenbacher, Park, Judd, & Correll, 2009). Specifically, the categorization task required participants to categorize portraits of women as Muslim and non-Muslim. In two consecutive blocks of 48 trials each, 24 portraits of women wearing traditionally tied Muslim headscarves (Muslim women) and 24 portraits of bareheaded women (non-Muslim women) were presented in random order. The 48 portraits presented in the categorization task were not shown as primes in the subsequent AMP. Again, however, the AMP did not reveal the expected preference for non-Muslim over Muslim women, $t(24) = 1.02$, $p = .32$, Cohen's $d = 0.20$. Accordingly, the absence of prejudice effects in the AMP did not seem to be a matter of category salience.

Overall Analysis

It has to be noted that, in consideration of the small sample sizes, the power to detect prejudice effects in each pilot

[2] Note that the pattern of crucial significant and nonsignificant findings remained the same when we reran analyses excluding the overall 22 participants who reported that relatives or close friends belonged to the out-group under investigation.

Table 1. Percentage of pleasant responses (*SD*) as a function of prime-type (out-group vs. in-group) and study (Studies 1–7), and tests for statistical significance of the prejudice effect

Study	out-group prime: *M* (*SD*)	in-group prime: *M* (*SD*)	*t*	*p*	*d*
Study 1 (*N* = 26): Turks versus Germans	51% (14%)	52% (13%)	0.87	.39	0.17
Study 2 (*N* = 27): Blacks versus Whites	55% (12%)	56% (11%)	0.56	.58	0.10
Study 3 (*N* = 24): Old versus Young	60% (14%)	59% (15%)	−1.09	.29	−0.22
Study 4 (*N* = 26): Muslims versus Non-Muslims	53% (16%)	56% (15%)	0.93	.36	0.17
Study 5 (*N* = 27): Arabs versus liked celebrities	49% (17%)	61% (13%)	3.11	.004	0.60
Study 6 (*N* = 27): Blacks versus Whites	59% (16%)	58% (14%)	−0.46	.65	−0.09
Study 7 (*N* = 25): Muslims versus Non-Muslims	54% (13%)	56% (14%)	1.02	.32	0.20

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

study was rather low. We therefore conducted an overall analysis across the seven pilot studies. Specifically, pleasant responses were submitted to a 2 (duration: $target_{short}$ vs. $target_{long}$) × 2 (prime valence: assumedly negatively valenced out-group vs. assumedly positively valenced in-group) × 2 (target valence: slightly negative vs. slightly positive) × 2 (group: first $block_{short}$ vs. first $block_{long}$) × 7 (study: Studies 1–7) analysis of variance with repeated measures on the first three factors. Note that we excluded the neutral prime trials from this analysis. This is because, due to our initial results, our research aim switched from investigating the processes underlying prejudice effects in the AMP (modeling approach requiring neutral prime trials) to finding prejudice effects with the AMP *at all*. For the latter research question, neutral prime trials are irrelevant.

Following Payne et al. (2010), the overall analysis of variance should reveal a main effect of prime valence and an interaction of prime valence and duration (i.e., stronger effects of prime valence for the $target_{short}$ duration than for the $target_{long}$ duration). Furthermore, albeit less interesting for the present research focus, a main effect of target valence is to be expected, and this factor should according to Payne et al. (2010) also interact with duration such that effects of target valence should be stronger for the $target_{long}$ duration than for the $target_{short}$ duration.

Table 2 presents all results in detail. Results were clear-cut and, by and large, replicated the findings of the separate analyses per study detailed in the ESM. In the overall analysis, the main effect of prime valence reached significance, $F(1, 168) = 6.32$, $p = .01$, $\eta_p^2 = .04$, indicating a preference for the in-group (M = 57% pleasant responses, SD = 13%) over the out-group (M = 54% pleasant responses, SD = 15%). However, this effect was qualified by the factor study, $F(6, 168) = 2.93$, $p = .01$, $\eta_p^2 = .10$, with the interaction of prime valence and study being driven by the data of Study 5. Remember that by contrasting terrorist-associated Arabs with liked celebrities, Study 5 used clearly valenced but naturally not opposing prime categories that may activate strong, socially less sensitive group-based attitudes in consideration of the currently strong fear of terrorist attacks. Furthermore, Study 5 confounded differences in social categories with strong differences in other variables (e.g., valence of emotional expression or attractiveness). When the data from Study 5 were excluded from analyses, the expected main effect of prime valence became non-significant, $F(1, 143) = .96$, $p = .33$, $\eta_p^2 = .01$. In contrast to Payne et al.'s (2010) hypotheses and findings, the duration variable did not interact with prime valence, independent of whether data from Study 5 were included in or excluded from analyses, $F < 1$ (for similar failures to replicate Payne et al.'s interaction effects, see Tobin & LeBel, 2014). The (non-) occurrence of prejudice effects was thus independent of the duration of target presentation. Further analyses confirmed that for both duration conditions ($target_{short}$ and $target_{long}$), prejudice effects were absent if data from Study 5 were excluded from analyses, ts < 1.

As in Payne et al.'s (2010) experiments, the main effect of target valence was significant, $F(1, 168) = 100.98$, $p < .001$, $\eta_p^2 = .38$, indicating more pleasant responses to slightly positive targets than to slightly negative targets. Furthermore, and in contrast to the separate analyses of each study's data, the expected interaction effect of target valence and duration reached significance in the overall analyses, $F(1, 174) = 6.36$, $p = .01$, $\eta_p^2 = .04$, with more pronounced effects of target valence for the $target_{long}$ duration than for the $target_{short}$ duration.

Interim Conclusion

To summarize, we failed to find significant prejudice effects in the AMP. Only if socially less sensitive attitudes were investigated and if the prime stimuli confounded differences in social categories with strong differences in other variables influenced by valence (see Study 5) did the expected AMP effects emerge. The present research thus suggests that the AMP may not be sensitive to prejudice against out-groups.

A number of alternative explanations to this conclusion could be ruled out. First, when using clearly valenced prime stimuli pre-validated in prior research, the AMP showed the expected evaluation effect (see results of Study 5). Thus, the absence of prejudice effects in six out of seven AMP studies

 © 2017 Hogrefe Publishing

Table 2. Results of the overall analysis of variance across data of the seven pilot studies

Effect	df_1	df_2	F	p	η_p^2
Group	1	168	0.26	.61	.00
Study	6	168	1.17	.32	.04
Group × Study	6	168	0.35	.91	.01
Duration	1	168	0.10	.75	.00
Duration × Group	1	168	0.26	.61	.00
Duration × Study	6	168	1.26	.28	.04
Duration × Group × Study	6	168	1.33	.25	.05
Prime	1	168	6.32	.01	.04
Prime × Group	1	168	1.10	.30	.01
Prime × Study	6	168	2.93	.01	.10
Prime × Group × Study	6	168	0.89	.50	.03
Target	1	168	100.98	< .001	.38
Target × Group	1	168	3.01	.08	.02
Target × Study	6	168	1.59	.15	.05
Target × Group × Study	6	168	0.65	.69	.02
Duration × Prime	1	168	0.02	.89	.00
Duration × Prime × Group	1	168	0.28	.60	.00
Duration × Prime × Study	6	168	0.74	.62	.03
Duration × Prime × Group × Study	6	168	0.61	.72	.02
Duration × Target	1	168	6.36	.01	.04
Duration × Target × Group	1	168	1.79	.18	.01
Duration × Target × Study	6	168	0.35	.91	.01
Duration × Target × Group × Study	6	168	0.93	.48	.03
Prime × Target	1	168	0.04	.84	.00
Prime × Target × Group	1	168	0.90	.34	.01
Prime × Target × Study	6	168	0.62	.71	.02
Prime × Target × Group × Study	6	168	0.59	.74	.02
Duration × Prime × Target	1	168	1.09	.30	.01
Duration × Prime × Target × Group	1	168	0.15	.70	.00
Duration × Prime × Target × Study	6	168	1.02	.41	.04
Duration × Prime × Target × Group × Study	6	168	2.14	.051	.07

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

cannot be explained in terms of specific procedural details preventing the occurrence of evaluation effects, in general. Second, increasing category salience in Study 7 (and thereby increasing the likelihood that the primes in the AMP were processed according to their social categories, see Olson & Fazio, 2003) did not result in the expected prejudice effects in the AMP. Thus, the absence of effects does not appear to be due to processing stimuli at an exemplar level instead of a category level. Such an explanation also would have conflicted with previous work by Gawronski et al. (2010) who showed that only for the EPT, but not for the AMP, effect sizes increased when participants' attention was directed to the social categories. Finally, results of the overall analyses suggest that the absence of prejudice effects does not simply result from too little power to detect prejudice effects (with the present overall sample size of 182, the power to detect a small effect of Cohen's d = 0.20 was .85).

Although we considered it an unlikely possibility, we also pursued the possibility that participants in our student sample may simply not hold prejudice against the out-groups under investigation. To evaluate this explanation of the present results, we conducted an additional study in which we examined prejudice as assessed by another prominent indirect measure, the IAT (Greenwald et al., 1998). Specifically, 20 participants who did not participate in the AMP studies, but were representative of our population of participants, completed two prejudice IATs: a Black-White IAT and a Muslim-non-Muslim IAT, with the order of IATs counterbalanced between participants. Importantly, the IATs used exactly the same stimuli as Study 6 and Studies 4 and 7, respectively. Both IATs produced large effects reflecting strong preferences for the in-groups over the out-groups with effect sizes of Cohen's $d \geq 1.50$ (for details, see the ESM). Given recent criticism of confounding influences on IAT effects (see Teige-Mocigemba et al., 2010),

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

we hesitate to interpret the absolute sizes of the IAT effects as pure measures of participants' attitudes. Specifically, it has been shown that IAT effects reflect not only the construct of interest (e.g., an attitude in an attitude IAT), but are also influenced by recoding effects, task-switching abilities, and block order effects, that, among other construct-unrelated variables, may boost IAT effects (e.g., Gawronski & De Houwer, 2014). These confounding influences notwithstanding, however, the direction of IAT effects may still be (and has often been) interpreted as a valid indicator of the direction of participants' preferences (e.g., Olson & Fazio, 2001; but see Blanton & Jaccard, 2006). After all, IAT effects have been found to comprise valid, construct-related variance (for an overview, see De Houwer et al., 2009). Accordingly, although the IAT's effect sizes may overestimate participants' preference for the in-group over the out-group (which is why we refrain from interpreting the absolute size of IAT effects), the direction of the present IAT effects suggests that our typical participant sample indeed prefers the in-group to the out-group (here, Whites over Blacks and non-Muslim women over Muslim women). In this regard, note that participants' preference for Turks over Germans as investigated in Study 1 was evidenced not only by IAT data, but also by evaluative-priming data (for details, see below).

All in all, the present research raises doubts about the AMP's suitability for the assessment of prejudice, and stands in contrast to recent conclusions about the robustness of AMP effects. It has to be noted, however, that these conclusions rely on the whole range of AMP findings in a multitude of attitude domains from which only few are concerned with prejudice or socially sensitive topics. In socially less sensitive domains (e.g., prejudice against members of opposing soccer teams), the AMP may indeed produce more robust effects than in socially more sensitive domains (e.g., prejudice against Turks).

In Search of Prejudice Effects: A Literature Review

To examine the robustness of AMP effects in socially more sensitive domains of prejudice, we reviewed past prejudice effects assessed by the AMP. We identified 19 papers reporting studies that implemented a variant of the AMP for the assessment of prejudice. In all studies, prejudice effects are to be expected and, accordingly, an indirect attitude measure such as the AMP should reveal such effects. Most studies used pictures of Black and White faces as primes, but others also used Arab versus European, Turkish versus German, old versus young, homosexual versus heterosexual, average weight versus underweight, or Jewish versus Christian as prime categories. Due to the many differences among the studies reported in the 19 papers with regard to, for instance, the prejudice domain of interest, the experimental setup, the procedural details of the AMP, and the country of data collection, conducting a meta-analysis appeared to be not informative at the present stage of evidence. Where possible, however, we computed Cohen's *d*. As can be seen in Table 3, there was a wide range of effect sizes across studies (Cohen's *d* ranging from −0.25 to 1.29).

On a more general level, many of the journal articles or university theses we reviewed found the expected prejudice effects (Cooley, Payne, & Phillips, 2014; Echabe, 2013; Gawronski & Ye, 2015; Payne et al., 2005, 2013; Payne, Krosnick et al., 2010; Pryor, Reeder, & Monroe, 2012). Many others, however, showed mixed evidence (Gawronski, Peters, Brochu, & Strack, 2008; Gawronski et al., 2010; Guinote, Willis, & Martellota, 2009; Payne, Burkley, & Stokes, 2008; Pinheiro & Wentura, 2012), no AMP effects (Proia, 2012), or even reversed AMP effects (Bar-Anan & Nosek, 2012; Lycke & Setterberg, 2011; Walsh, 2013).[3] Mixed and reverse AMP effects were shown in samples from different countries, and with different prime categories. In line with our own findings reported above, the literature review thus indicates that AMP effects in socially sensitive prejudice domains are not very robust. Considering the well-known fact that null findings are hard to publish (Rosenthal, 1979) and given the current academic incentive structure (e.g., Nosek & Bar-Anan, 2012), the literature review may even overestimate the evidence of prejudice effects assessed by the AMP. It therefore appears to be a question worthy of further study whether or not the AMP is able to detect an evaluative out-group bias when it is present. In this regard, a preregistered study investigating this question guarantees by the nature of the preregistration procedure that its results are free of selection and publication biases.

Preregistered Study

Given that our pilot studies reported above investigated different domains of prejudice, used small sample sizes, and employed a modified AMP version, we ran a final, high-powered AMP study to provide a more conclusive answer to the question of whether the AMP does reveal existing prejudice effects in Germany or not. The results of such a study could be an important starting point for further research on possible variables that may moderate or mediate the occurrence of evaluation effects in the AMP.

[3] Please note that three studies were not listed above because they did not report AMP effects (Imhoff & Banse, 2009; Lundberg & Payne, 2014) or were not interpretable for the present purpose (Todd & Burgmer, 2013).

 © 2017 Hogrefe Publishing

Table 3. Effect sizes (Cohen's *d*) of AMP studies on prejudice

Paper	Study	*N*	Cohen's *d*
Bar-Anan and Nosek (2012)	Experiment 1a	354	−0.10
	Experiment 3	702	−0.18
Cooley et al. (2014)	Experiment 1	115	1.00
	Experiment 2	147	0.64
	Experiment 3	82	0.39
Echabe (2013)	Experiment 1	42	1.23
	Experiment 3	50	1.29
Gawronski et al. (2008)	Experiment 2	77	0.05
	Experiment 3	95	0.33
Gawronski et al. (2010)	Pilot study – race effect	76	0.21
	Pilot study – age effect	76	0.27
	Experiment – race effect	97	0.20
	Experiment – age effect	97	0.20
Payne et al. (2005)	Experiment 6	35	0.50
Payne, Burkley, et al. (2008)	Experiment 1	75	0.26
	Experiment 2	48	−0.08
	Experiment 4	71	0.31
Payne, Krosnick, et al. (2010)	Experiment 1	1,056	0.35
	Experiment 2	1,933	0.45
	Experiment 3	1,424	0.20
Payne et al. (2013)	Experiment 1	288	0.24
Proia (2012)	Experiment 1	82	−0.14
	Experiment 2	57	−0.25
Pryor et al. (2012)	Experiment 1	77	0.66
	Experiment 3	176	0.15

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Specifically, we aimed to investigate prejudice against Turks, the biggest minority in Germany, by means of the standard AMP as introduced by Payne et al. (2005; see also Payne & Lundberg, 2014). Based on our pilot studies and the literature review, but in contrast to recent conclusions about the robustness of AMP effects, we had doubts that the AMP would reveal prejudice effects, although prejudice is expected to be prevalent. For instance, different studies conducted in Germany using indirect measures (e.g., EPT, IAT, Extrinsic Affective Simon Task, stereotype priming), direct self-report measures, and behavioral data suggest that Germans are prejudiced against Turks (e.g., Asbrock, 2010; Degner & Wentura, 2008, 2011; Florack, Scarabis, & Bless, 2001; Gawronski, 2002; Gawronski, Geschke & Banse, 2003; Gschwendner, Hofmann, & Schmitt, 2006; Hofmann, Gschwendner, Castelli, & Schmitt, 2008; Kahraman & Knoblich, 2000; Klauer, Schmitz, Teige-Mocigemba, & Voss, 2010; Klink & Wagner, 1999; Neumann & Seibt, 2001; Pettigrew, 1998; Wagner, van Dick, & Zick, 2000; but see Degner, Wentura, Gniewosz, & Noack, 2007). Importantly, in most of these studies, participants have been students indicating that prejudice against Turks is prevalent in the population of German students.

To get an empirically grounded estimate on prejudice against Turks in our population of Freiburg participants, we analyzed data of two further independent studies using indirect measures, one EPT study and one IAT study. Importantly, both studies relied mainly on University-of-Freiburg students (i.e., the samples were representative of our population of participants) and employed the same prime stimuli as those we used in the preregistered AMP study (see below). Results of the two studies should therefore reveal whether and to what extent our participants can be expected to hold prejudice against Turks. As detailed in the ESM, we found a clear preference for Germans over Turks in both indirect measures (Cohen's d_{EPT} = 0.53; Cohen's d_{IAT} = 2.10). Accordingly, the AMP should also reflect this evaluative out-group bias if it is an indirect measure of prejudice.

Method

Participants

Different approaches have been proposed to determine a study's sample size on a priori grounds. For instance,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

sample size estimations may be based on effect sizes that (a) could be considered as meaningful in the prejudice context, that (b) have been found in one's own studies (and, thus, may be expected for the population of participants), or that (c) different measures typically reveal in the prejudice domain of interest. We followed these and further approaches, and determined sample size by complying with the most conservative result. The most conservative approach turned out to be taking an effect size that, by convention, marks the lower bound of small effects (i.e., Cohen's d = 0.20; cf. Lakens & Evers, 2014). With a power of .90 and a Type 1 error probability of .05, we needed 216 participants to detect an existing prejudice effect by means of a one-tailed t test. As a second approach, we calculated a 95% confidence interval for the lower effect size of the two indirect measures on prejudice against Turks that our student sample worked through (i.e., Cohen's d_{EPT} = 0.53 < Cohen's d_{IAT} = 2.10; see ESM for details), using a script provided by Wuensch (2012). With the lower bound of this interval, Cohen's d = 0.29, we would only need 102 participants to detect an existing prejudice effect with a power of .90 (power would be .995 with N = 216). Finally, we could have chosen an approach recommended by Simonsohn (2015), according to which the sample size of the original prejudice AMP (Payne et al., 2005, Experiment 6) has to be multiplied by 2.5. This would have resulted in an N of 138, which is considerably below the number of participants we decided to aim for.

Given our decision to comply with the most conservative approach, we aimed at recruiting complete data sets of 216 German participants (with mother tongue German) who were mainly students of the University of Freiburg. Only participants aged between 18 and 45 years were allowed to participate to ensure task comprehension and sufficient task performance. For participation, participants received €1.50. Thirty-eight participants had to be excluded from data analyses due to a priori defined exclusion criteria[4] and were replaced by newly gathered participants to ensure the final sample size of 216 for the analyses. It should be noted, however, that the significance pattern of the results reported below did not change if data analyses included all participants. Of the 216 German participants (124 female, 91 male, and 1 undeclared; mean age = 23, SD = 3.85, range 18–45 years) whose data entered the analyses, 26 indicated to have Turkish relatives or close Turkish friends (189 participants denied, 1 undeclared).

Procedure

In individual sessions of about 15 min, participants first completed the Turks versus Germans AMP. Subsequently, participants were asked to provide some personal data (i.e., age, gender, handedness, major, familiarity with Chinese language, mother tongue, and whether they have Turkish relatives or close Turkish friends). Finally, participants were thanked, compensated for participation, and debriefed.

The AMP

Most importantly, the preregistered study used a standard AMP to exclude the possibility that the block-wise within manipulation of duration of target presentation may have influenced results of the pilot AMP studies. Furthermore, we only presented portraits of Turks and Germans as primes, but not of neutral primes because these were irrelevant for the present research question. Participants were seated at a distance of approximately 55 cm from the computer screen.

Materials

Prime and target stimuli had a size of 256 × 256 pixels (subtending between approximately 7 cm horizontally and 7 cm vertically on the screen) and were shown on a white background on a 58 cm LCD TFT monitor (LG Flatron W2363D) with a refresh rate of 120 Hz.

Primes

Prime stimuli were six portraits representing Turks and six portraits representing Germans. We selected these stimuli from a stimulus pool provided by Singmann, Kellen, and Klauer (2013), who also made the corresponding stimulus ratings of an online study available to us. Pre-ratings comprised ratings of typicality (possible range: −6 = *very*

[4] Following the preregistered exclusion criteria, six participants were excluded because they indicated that they were able to read Chinese characters, two participants were aged below 18, and 8 participants indicated a mother tongue other than German and were therefore excluded.

In reaction to issues that arose during data collection, 22 further participants were excluded from data analyses due to incomplete data, unwanted interruptions of the test session, prior participation in related experiments, and technical problems: Specifically, for one participant, demographic variables were missing, and another participant did not read instructions and interrupted the experiment to ask the experimenter about what to do in the AMP; both participants were excluded. Seven further participants were excluded because they participated beforehand in related experiments that we had a priori defined as too closely related to the current experiment for such participants to count as "naïve" participants. These experiments comprised the pilot studies reported above and experiments that used different kinds of prejudice reduction interventions. Finally, data of 13 participants were excluded due to technical problems that led to minimal unintended deviations between the procedural details of the preregistered and the actually presented AMP procedure: Either a single prime stimulus was presented 13 times instead of 12 times (and, in return, another prime stimulus only 11 times instead of 12 times) or a single Chinese character was presented twice instead of once (and, in return, another Chinese character was not presented) across all AMP trials. Note that there was not a single participant who had to be excluded because his/her use of one response key was a "far-out" value (Tukey, 1977) in the distribution of the total sample (this was preregistered as another exclusion criterion; see Payne & Lundberg, 2014).

 © 2017 Hogrefe Publishing

typical for the other race, 6 = *very typical for the own race*) and of valence of emotional expression (1 = *negative*, 7 = *positive*). Ratings of attractiveness (1 = *very unattractive*, 5 = *very attractive*) were collected in our laboratories (N = 40; 20 female, 20 male; mean age = 25 years). Stimuli were selected such that they were matched with regard to typicality of their respective social group (M_{Turks} = 4.32, SD_{Turks} = .53; $M_{Germans}$ = 4.28, $SD_{Germans}$ = .54; t[10] = .14, p = .89), valence of emotional expression (M_{Turks} = 4.33, SD_{Turks} = .65; $M_{Germans}$ = 4.34, $SD_{Germans}$ = .59; t[10] = −.02, p = .98), and attractiveness (M_{Turks} = 2.67, SD_{Turks} = .27; $M_{Germans}$ = 2.80, $SD_{Germans}$ = .34; t[10] = −.77, p = .46) to control for variables that may confound differences in social categories. Importantly and as reported above, an IAT and an EPT using exactly the same prime stimuli have been found to reveal robust prejudice effects, confirming the suitability of the preselected stimuli (see ESM for further detail).[5]

Targets

As targets, we used the same 144 target stimuli that we had also used in Studies 2–7. According to our reanalyses of prior AMP data (see above), these stimuli belong to the most moderately valenced Chinese characters of Keith Payne's target pool that is typically used for AMP studies.

AMP Trials

An AMP trial started with a prime picture presented for 75 ms in the center of the screen, followed by a blank screen presented for 125 ms. Then, a target appeared that was masked after 100 ms by a pattern mask consisting of black and white noise. The pattern mask remained on screen until participants responded. Responses were given by pressing the [Q] for visually more unpleasant responses and the [P] for visually more pleasant responses on a standard German QWERTZ keyboard. The labels "unpleasant" and "pleasant" were presented in the corresponding corners of the screen throughout the experiment.

Sequence of AMP Trials

Participants first completed a practice block, followed by two test blocks. The practice block consisted of four trials with one Turk stimulus and one German stimulus being paired with a slightly pleasant target stimulus, and another Turk stimulus and German stimulus being paired with a slightly unpleasant target stimulus. Primes and targets were paired in an otherwise random order. All stimuli of the practice block (two Turk stimuli, two German stimuli, and four target stimuli) were reserved for the practice block only.

The test blocks comprised 72 trials each, with each of the 12 prime pictures being presented six times per block and each of the 144 target pictures presented only once throughout the two test blocks. Per test block, each prime stimulus was paired three times with a slightly negative target and three times with a slightly positive target, and prime-target pairs were presented in an otherwise random order.

Results

For the sake of comparability with the results of the pilot studies, we submitted pleasant responses to a 2 (prime valence: assumedly negatively valenced Turks vs. assumedly positively valenced Germans) × 2 (target valence: slightly negative vs. slightly positive) analysis of variance with repeated measures on both factors. We were mostly interested in the main effect of prime valence. As in the pilot studies, however, and excluding the possibility that participants provided random responses, we expected to find a main effect of target valence, with more pleasant responses for slightly pleasant as compared to slightly unpleasant targets.

Indeed, the main effect of target valence reached significance, F(1, 215) = 76.05, p < .001, η_p^2 = .26, replicating the pattern of the pilot studies. Most importantly, however, the

[5] An anonymous reviewer suspected that our efforts to equate the stimulus sets with regard to attractiveness might result in the selection of Turks who are more attractive than the selected Germans. This is because participants' prejudice may have biased attractiveness ratings leading to generally worse evaluations of Turks on positive dimensions such as attractiveness. Any null effect in the AMP may, thus, reflect diametrically opposed influences of prejudice and biased stimulus selection that cancel each other out.

From our perspective, two findings contradict this assumption. First, as detailed in the main text and the ESM, we found significant prejudice effects in an evaluative-priming task and an IAT that used exactly the same stimuli as the AMP in the preregistered study. If the portraits of Turks were more attractive than those of Germans, this should have diminished prejudice effects in the EPT and the IAT as well.

Second, we followed the reviewer's advice and collected attractiveness ratings of an additional Turkish sample (N = 40; 20 female, 20 male; mean age = 24 years), assuming that Turkish participants can be expected to give more unbiased attractiveness judgments of portrayed in-group members. Replicating the findings of the German sample, Turkish participants rated the preselected portraits of Turks and of Germans as similarly attractive (M_{Turks} = 2.39, SD_{Turks} = .23; $M_{Germans}$ = 2.58, $SD_{Germans}$ = .40; t[10] = −.995, p = .34). Descriptively, the selected portraits of Turks were even rated as less attractive than the selected portraits of Germans. This should, if anything, add to, rather than run counter to, prejudice effects in the AMP. We also submitted attractiveness ratings to a 2 (sample: German participants vs. Turkish participants) × 2 (prime: portrait of a German vs. portrait of a Turk) analysis of variance with repeated measures on the second factor. This analysis revealed a significant main effect of sample, F(1, 10) = 41.38, p < .001, with the Turkish participants giving slightly lower attractiveness ratings overall than the German participants. Importantly, however, neither the main effect of prime, F(1, 10) = 0.82, p = .39, nor the interaction of sample and prime, F(1, 10) = 0.41, p = .54, reached significance.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

main effect of prime valence was also significant, $F(1, 215) = 6.00$, $p = .015$, $\eta_p^2 = .03$, indicating a preference for Germans ($M = 55.45\%$ pleasant responses, $SD = 13.13\%$) over Turks ($M = 53.68\%$ pleasant responses, $SD = 13.50\%$). This AMP effect, however, was very small, $M = 1.77\%$, $SD = 10.61$, $t(215) = 2.45$, $p = .008$ (one-tailed), Cohen's $d = 0.17$. The interaction of prime valence and target valence was not significant, $F(1, 215) = .003$, $p = .95$, $\eta_p^2 = .00$.

In order to exclude the possibility that the prejudice effect may have been depressed by participants who can be assumed to be positively biased toward Turks on a priori grounds, we reran analyses after having excluded the 26 participants who reported to have Turkish relatives or close Turkish friends and one further participant who did not respond to this item. As in the pilot studies, the reanalysis revealed the same pattern of significances as the analysis of the whole sample: Again, there was a small, albeit significant, AMP effect of $M = 1.68\%$, $SD = 10.44$, $t(188) = 2.21$, $p = .015$ (one-tailed), Cohen's $d = 0.16$.

Upon the editor's request, we tested as an exploratory analysis whether the very small effect goes back to a few participants with particularly strong effects (cf. Bar-Anan & Nosek, 2012). There were indeed six participants with outlier values in the distribution of the individual AMP effects (Tukey, 1977). Specifically, with values larger than or equal to 33%, AMP effects of five participants were above the third quartile plus 1.5 times the total sample's interquartile range of the AMP effect, while, with a value of −35%, the AMP effect of only one participant was below the first quartile minus 1.5 times the total sample's interquartile range of the AMP effect. Not surprisingly, the AMP effect was not normally distributed as revealed by the Shapiro-Wilk test, $W = .96$, $p < .001$. Excluding the six participants with outlier values led to a normal distribution of AMP effects, $W = .99$, $p = .23$. Furthermore, although the AMP effect decreased, it still reached significance in the reduced sample, $M = 1.09\%$, $SD = 8.84$, $t(209) = 1.79$, $p = .04$ (one-tailed), Cohen's $d = 0.12$.

Finally, as an exploratory analysis that we had preregistered, we conducted Bayesian analyses in order to gain information about the strength of evidence for the presence (alternative hypothesis) or absence (null hypothesis) of a prejudice effect in the AMP. Bayesian analyses were performed using the distributions recommended by Rouder, Speckman, Sun, Morey, and Iverson (2009; Cauchy prior width of .707). The Bayes factor (BF) evidence for the alternative hypothesis in a one-tailed t test amounted to 2.80 for the deviation of the AMP effect from zero. According to conventional classifications (e.g., Wetzels & Wagenmakers, 2012), the BF evidence in favor of the alternative hypothesis (i.e., a prejudice effect in the AMP) was thus anecdotal (including BFs between 1 and 3).

General Discussion

The Affect Misattribution Procedure has been forwarded as one of the most promising alternatives to the Implicit Association Test and the evaluative-priming task for measuring attitudes such as prejudice indirectly. In the present research, we investigated whether the AMP is indeed able to detect an evaluative out-group bias. In contrast to recent conclusions about the robustness and large effect sizes of AMP effects (Payne et al., 2005; Payne & Lundberg, 2014), six out of seven pilot studies had indicated that participants did not show any prejudice effects in the AMP. Yet, these pilot studies were not fully conclusive with regard to our research question because they investigated different domains of prejudice, used small sample sizes, and employed a modified AMP version. In a preregistered, high-powered AMP study, we therefore examined whether the standard AMP as introduced by Payne and colleagues (2005; see also Payne & Lundberg, 2014) does reveal prejudice against the biggest minority in Germany: Turks.

Across different studies conducted in Germany, indirect measures other than the AMP (e.g., EPT, IAT, Extrinsic Affective Simon Task, stereotype priming) converge on the finding that Germans (including students) are prejudiced against Turks (see above for further details). The AMP should therefore reflect this evaluative out-group bias if it is an indirect measure of attitudes. In the present preregistered study, the AMP indeed revealed a significant prejudice effect. However, with a Cohen's d of 0.17, this effect was much smaller than what we (and other researchers) had found with other (indirect) measures (e.g., for the EPT: Cohen's $d = 0.53$; for the IAT: Cohen's $d = 2.10$). In fact, the prejudice effect in the AMP was so small that it even fell short of the lower bound for a small effect (i.e., Cohen's $d = 0.20$), and, in terms of Bayesian analyses, the evidence in favor of a prejudice effect was only anecdotal.

Taken together, the present findings conflict with recent conclusions about the robustness and large effect sizes of AMP effects. Particularly in socially more sensitive domains, such as prejudice against out-group members, the AMP appears to reveal, if anything, only small effects as was also indicated by a literature review reported above.[6] With regard to prejudice against Turks, the very small effect

[6] Note, however, that the present findings do not exclude the possibility that an AMP effect correlates with external criteria in a meaningful way, irrespective of its overall mean.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

 © 2017 Hogrefe Publishing

found in the present preregistered study is particularly surprising because data collection of this study coincided with crucial events (e.g., terrorist attacks conducted by members of the Islamic State in the neighboring countries France and Belgium) and public debates (e.g., media coverage of refugees and Muslims in Germany) that reinforced more negative perceptions of Arabs and Muslims (e.g., as possibly threatening) and, perhaps, of Turks as well. Accordingly, it may be expected that prejudice against Turks is currently even augmented, which was, however, not reflected in the size of the AMP effect, thereby questioning the AMP's sensitivity to prejudice effects.

Clearly, the present findings raise the question of how to account for the AMP's weak sensitivity to prejudice effects. One possibility is that participants may deliberately distort their responses when it comes to the assessment of socially sensitive attitudes. After all, it has been shown that the AMP can be faked easily (Teige-Mocigemba et al., 2016). For instance, participants may effectively prevent the primes from influencing responses to the targets (e.g., by actively contrasting target judgments away from the evaluative implications of the primes, see Krieglmeyer & Sherman, 2012). This may also lead to more accurate target evaluations, which have been shown to reduce the size of AMP effects (e.g., see Eder & Deutsch, 2015; Teige-Mocigemba et al., 2016). Alternatively, in socially sensitive domains, participants might be more hesitant to intentionally rate the prime instead of the target, when working through the AMP (see Bar-Anan & Nosek, 2012). Note that this explanation may also account for the significant AMP effect in Pilot Study 5 that used not only clearly valenced stimulus materials, but also examined prejudice in a socially less sensitive domain (namely, preference for liked celebrities over terrorist-associated Arabs). In other words, to the extent to which participants are not motivated to hide their attitudes, evaluation effects may be more likely to increase the size of AMP effects.

Undoubtedly, future research is needed to examine the impact of the just proposed and further explanations. We believe that such research is worth the effort because it may help (a) to improve an indirect measure that has become one of the most promising candidates among indirect measures and (b) to better understand what variables influence the processes assumed to underlie AMP effects (e.g., see Gawronski & Ye, 2014). For instance, if it turned out that strategic control can account for the AMP's small effect sizes in socially sensitive domains, changing some procedural details such as, for example, implementation of a response deadline may reduce participants' control efforts considerably. Moreover, there might be variables that increase the likelihood of the misattribution process to operate, and, thereby, may enhance the AMP's sensitivity to spontaneous evaluations, even in socially sensitive domains. For instance, and as implied by the "inkblot" metaphor used by Payne and colleagues (2005), the targets have to be ambiguous enough to permit the misattribution of prime evaluations. Beyond that, however, it might be hypothesized that prime-target similarity also influences the likelihood of the misattribution process. Imagine, for instance, that we had employed a prejudice AMP that used more human, face-like shapes as targets such as the targets of the Stereotype Misperception Task of Krieglmeyer and Sherman (2012). It appears to be plausible that it would be easier to misattribute a spontaneous evaluation of a face (e.g., a portrait of a Turk) to these more human, face-like targets than to the nonhuman Chinese characters that are typically used in a standard AMP. Put differently, besides the targets' ambiguity, the applicability of prime evaluations to target evaluations may be important for the operation of the misattribution process and, thus, for the size of AMP effects (cf. Higgins, Rholes, & Jones, 1977).

Conclusion

The main purpose of the present research was to examine whether the AMP is able to detect an evaluative out-group bias when it is present. In contrast to recent conclusions about the robustness of AMP effects, several pilot studies did not find significant out-group bias and a preregistered, high-powered study revealed an effect that was smaller than what is conventionally labeled as a small effect (i.e., Cohen's d was smaller than 0.20). The present findings thereby suggest poor sensitivity of the AMP as a measure of prejudice. Given that AMP effects appear to be more robust in socially less sensitive domains, we discussed possible reasons for the AMP's weak sensitivity to evaluations in socially sensitive domains and sketched possible lines of future research. We hope that the present work will inspire future research on possible variables that may moderate or mediate the occurrence of evaluation effects in the AMP.

Acknowledgments

The research reported in this paper was supported by the Anneliese Maier Research Award of the Alexander von Humboldt Foundation to Jeffrey W. Sherman.

Electronic Supplementary Materials

The electronic supplementary material is available with the online version of the article at http://dx.doi.org/10.1027/1618-3169/a000364. Raw data can be found at https://osf.io/vcz4a

ESM 1. Pilot Studies (doc).
Report details on each pilot studies' sample, method, and results.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

References

Asbrock, F. (2010). Stereotypes of social groups in Germany in terms of warmth and competence. *Social Psychology, 41*, 76–81. doi: 10.1027/1864-9335/a000011

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83*, 380–393. doi: 10.1037/0022-3514.83.2.380

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7*, 543–554. doi: 10.1177/1745691612459060

Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the affective misattribution procedure. *Personality and Social Psychology Bulletin, 38*, 1194–1208. doi: 10.1177/0146167212446835

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *The American Psychologist, 61*, 27–41. doi: 10.1037/0003-066X.61.1.27

Cameron, C. D., Brown-Iannuzzi, J., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behaviors and explicit attitudes. *Personality and Social Psychology Review, 16*, 330–350. doi: 10.1177/1088868312440047

Cooley, E., Payne, B. K., & Phillips, K. J. (2014). Implicit bias and the illusion of conscious ill will. *Social Psychological and Personality Science, 5*, 500–507. doi: 10.1177/1948550613506123

Craig, B. M., Lipp, O. V., & Mallan, K. M. (2014). Emotional expression is preferentially processed in implicit evaluations of faces varying in race or age. *Emotion, 14*, 865–877. doi: 10.1037/a0037270

Deffenbacher, D. M., Park, B., Judd, C. M., & Correll, J. (2009). Category boundaries can be accentuated without increasing intergroup bias. *Group Processes & Intergroup Relationships, 12*, 175–193. doi: 10.1177/1368430208101055

Degner, J., & Wentura, D. (2008). The extrinsic affective Simon task as an instrument for indirect assessment of prejudice. *European Journal of Social Psychology, 38*, 1033–1043. doi: 10.1002/ejsp.536

Degner, J., & Wentura, D. (2011). Types of automatically activated prejudice: Assessing possessor-versus other-relevant valence in the evaluative priming task. *Social Cognition, 29*, 182–209. doi: 10.1521/soco.2011.29.2.182

Degner, J., Wentura, D., Gniewosz, B., & Noack, P. (2007). Hostility-related prejudice against Turks in adolescents: Masked affective priming allows for a differentiation of automatic prejudice. *Basic and Applied Social Psychology, 29*, 245–256. doi: 10.1080/01973530701503150

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*, 347–368. doi: 10.1037/a0014211

Echabe, A. E. (2013). Relationship between implicit and explicit measures of attitudes: The impact of application conditions. *Europe's Journal of Psychology, 9*, 231–245. doi: 10.5964/ejop.v9i2.544

Eder, A. B., & Deutsch, R. (2015). Watch the target! Effects in the affective misattribution procedure become weaker (but not eliminated) when participants are motivated to provide accurate responses to the target. *Frontiers in Psychology, 6*, 1442. doi: 10.3389/fpsyg.2015.01442

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013–1027. doi: 10.1037/0022-3514.69.6.1013

Florack, A., Scarabis, M., & Bless, H. (2001). Der Einfluss wahrgenommener Bedrohung auf die Nutzung automatischer Assoziationen bei der Personenbeurteilung [The influence of perceived threat on the use of automatic associations in person judgments]. *Zeitschrift für Sozialpsychologie, 32*, 249–259. doi: 10.1024/0044-3514.32.4.249

Gawronski, B. (2002). What does the Implicit Association Test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology, 49*, 171–180. doi: 10.1026/1618-3169.49.3.171

Gawronski, B., Cunningham, W. A., LeBel, E. P., & Deutsch, R. (2010). Attentional influences on affective priming: Does categorisation influence spontaneous evaluation of multiply categorisable objects? *Cognition and Emotion, 24*, 1008–1025. doi: 10.1080/02699930903112712

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed, pp. 283–310). New York, NY: Cambridge University Press.

Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression formation: Associations influence the construal of individuating information. *European Journal of Social Psychology, 33*, 573–589. doi: 10.1002/ejsp.166

Gawronski, B., & Payne, B. K. (2010). *Handbook of implicit social cognition: Measurement, theory, and applications*. New York, NY: Guilford Press.

Gawronski, B., Peters, K. R., Brochu, P. M., & Strack, F. (2008). Understanding the relations between different forms of racial prejudice: A cognitive consistency perspective. *Personality and Social Psychology Bulletin, 34*, 648–665. doi: 10.1177/0146167207313729

Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency. Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology, 40*, 335–542. doi: 10.1016/j.jesp.2003.10.005

Gawronski, B., & Ye, Y. (2014). What drives priming effects in the affect misattribution procedure? *Personality and Social Psychology Bulletin, 40*, 3–15. doi: 10.1177/0146167213502548

Gawronski, B., & Ye, Y. (2015). Prevention of intention invention in the affect misattribution procedure. *Social Psychological and Personality Science, 6*, 101–108. doi: 10.1177/1948550614543029

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology, 108*, 553–561. doi: 10.1037/pspa0000016

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480. doi: 10.1037/0022-3514.74.6.1464

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17–41. doi: 10.1037/a0015575

Gschwendner, T., Hofmann, W., & Schmitt, M. (2006). Synergistic moderator effects of situation and person factors of awareness and adjustment on the consistency of implicit and explicit attitudes. *Journal of Individual Differences, 27*, 47–56. doi: 10.1027/1614-0001.27.1.47

Guinote, A., Willis, G. B., & Martellota, C. (2009). Social power increases implicit prejudice. *Journal of Experimental Social Psychology, 46*, 299–307. doi: 10.1016/j.jesp.2009.11.012

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13*, 141–154. doi: 10.1016/s0022-1031(77)80007-3

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

 © 2017 Hogrefe Publishing

Hofmann, W., Gschwendner, T., Castelli, L., & Schmitt, M. (2008). Implicit and explicit attitudes and interracial interaction: The moderating role of situationally available control resources. *Group Processes & Intergroup Relations, 11*, 69–87. doi: 10.1177/1368430207084847

Hofmann, W., Gschwendner, T., Nosek, B. A., & Schmitt, M. (2005). What moderates implicit-explicit consistency? *European Review of Social Psychology, 16*, 335–390. doi: 10.1080/10463280500443228

Imhoff, R., & Banse, R. (2009). Ongoing victim suffering increases prejudice. *Psychological Science, 20*, 1443–1447. doi: 10.1111/j.1467-9280.2009.02457.x

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76*, 349–364. doi: 10.1037/h0031617

Kahraman, B., & Knoblich, G. (2000). "Stechen statt sprechen" : Valenz und Aktivierbarkeit von Stereotypen über Türken [Valence and activation of stereotypes about Turks]. *Zeitschrift für Sozialpsychologie, 31*, 31–34. doi: 10.1024/0044-3514.31.1.31

Klauer, K. C., Schmitz, F., Teige-Mocigemba, S., & Voss, A. (2010). Understanding the role of executive control in the Implicit Association Test: Why flexible people have small IAT effects. *The Quarterly Journal of Experimental Psychology, 63*, 595–619. doi: 10.1080/17470210903076826

Klink, A., & Wagner, U. (1999). Discrimination against ethnic minorities in Germany: Going back to the field. *Journal of Applied Social Psychology, 29*, 402–423. doi: 10.1111/j.1559-1816.1999.tb01394.x

Krieglmeyer, R., & Sherman, J. W. (2012). Disentangling stereotype activation and stereotype application in the Stereotype Misperception Task. *Journal of Personality and Social Psychology, 103*, 205–224. doi: 10.1037/a0028764

Lakens, D., & Evers, E. R. K. (2014). Sailing from the seas of chaos into the corridors of stability: Practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science, 9*, 278–292. doi: 10.1177/1745691614528520

Lundberg, K. B., & Payne, B. K. (2014). Decisions among the undecided: Implicit attitudes predict future voting behavior of undecided voters. *PLoS One, 9*, e85680. doi: 10.1371/journal.pone.0085680

Lycke, J., & Setterberg, P. (2011). *Effects of allocation of attention on initial facial race discrimination*, Unpublished bachelor's thesis. Stockholm, Sweden: Lund University.

Neumann, R., & Seibt, B. (2001). The structure of prejudice: Associative strength as a determinant of stereotype endorsement. *European Journal of Social Psychology, 31*, 609–620. doi: 10.1002/ejsp.69

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259. doi: 10.1037/0033-295X.84.3.231

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry, 23*, 217–243. doi: 10.1080/1047840X.2012.692215

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology, 45*, 137–141. doi: 10.1027/1864-9335/a000192

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413–417. doi: 10.1111/1467-9280.00376

Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of racial prejudice: What are we measuring? *Psychological Science, 14*, 636–639. doi: 10.1046/j.0956-7976.2003.psci_1477.x

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716. doi: 10.1126/science.aac4716

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology, 105*, 171–192. doi: 10.1037/a0032734

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology, 108*, 562–571. doi: 10.1037/pspa0000023

Payne, B. K., Brown-Iannuzzi, J., Burkley, M., Arbuckle, N. L., Cooley, E., Cameron, C. D., & Lundberg, K. B. (2013). Intention invention and the affect misattribution procedure: Reply to Bar-Anan and Nosek (2012). *Personality and Social Psychology Bulletin, 39*, 375–386. doi: 10.1177/0146167212475225

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit tests diverge? The role of structural fit. *Journal of Personality and Social Psychology, 94*, 16–31. doi: 10.1037/0022-3514.94.1.16

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*, 277–293. doi: 10.1037/0022-3514.89.3.277

Payne, B. K., Govorun, O., & Arbuckle, N. L. (2008). Automatic attitudes and alcohol: Does implicit liking predict drinking? *Cognition and Emotion, 22*, 238–271. doi: 10.1080/02699930701357394

Payne, B. K., Hall, D., Cameron, C. D., & Bishara, A. J. (2010). A process model of affect misattribution. *Personality and Social Psychological Bulletin, 36*, 1397–1408. doi: 10.1177/0146167210383440

Payne, B. K., Krosnick, J. A., Pasek, J., Lelkes, Y., Akhtar, O., & Tompson, P. (2010). Implicit and explicit prejudice in the 2008 American presidential election. *Journal of Experimental Social Psychology, 46*, 367–374. doi: 10.1016/j.jesp.2009.11.001

Payne, B. K., & Lundberg, K. B. (2014). The Affect Misattribution Procedure: Ten years of evidence on reliability, validity, and mechanisms. *Social and Personality Psychology Compass, 8*, 672–686. doi: 10.1111/spc3.12148

Pettigrew, T. F. (1998). Reactions towards the new minorities of Western Europe. *Annual Review of Sociology, 24*, 77–103. doi: 10.1146/annurev.soc.24.1.77

Pinheiro, M. C., & Wentura, D. (2012). Automatically activated facets of ageism: Masked evaluative priming allows for a differentiation of age-related prejudice. *European Journal of Social Psychology, 42*, 852–863. doi: 10.1002/ejsp.1912

Proia, A. G. (2012). *The role of affect in the control of attention to monoracial and racially ambiguous faces*. Unpublished bachelor's thesis. Williamsburg, VA.

Pryor, J. B., Reeder, G. D., & Monroe, A. E. (2012). The infection of bad company: Stigma by association. *Journal of Personality and Social Psychology, 102*, 224–241. doi: 10.1037/a0026270

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638–641. doi: 10.1037/0033-2909.86.3.638

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237. doi: 10.3758/PBR.16.2.225

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*, 559–569. doi: 10.1177/0956797614567341

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Singmann, H., Kellen, D., & Klauer, K. C. (2013). Investigating the other-race effect of Germans toward Turks and Arabs using multinomial processing tree models. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1330–1335). Austin, TX: Cognitive Science Society.

Teige-Mocigemba, S., & Klauer, K. C. (2013). On the controllability of evaluative-priming effects: Some limits that are none. *Cognition and Emotion, 27*, 632–657. doi: 10.1080/02699931.2012.732041

Teige-Mocigemba, S., & Klauer, K. C. (2015). Implicit association test. In J. D. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (2nd ed., Vol. 11, pp. 703–708). Oxford, UK: Elsevier.

Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). Practical guide to implicit association test and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. S. 117–139). New York, NY: Guilford Press.

Teige-Mocigemba, S., Penzl, B., Becker, M., Henn, L., & Klauer, K. C. (2016). Controlling the "Uncontrollable": Faking effects on the affect misattribution procedure. *Cognition and Emotion, 30*, 1470–1484. doi: 10.1080/02699931.2015.1070793

Tobin, C., & LeBel, E. P. (2014, June 27). *Replication difficulties of Payne, Hall, Cameron, and Bishara's (2010) multinomial processing tree model of the affect misattribution procedure*. Available at SSRN: http://ssrn.com/abstract=2459994

Todd, A. R., & Burgmer, P. (2013). Perspective taking and automatic intergroup evaluation change: Testing an associative self-anchoring account. *Journal of Personality and Social Psychology, 104*, 786–802. doi: 10.1037/a0031999

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Weasley.

Wagner, U., van Dick, R., & Zick, A. (2000). Sozialpsychologische Analysen und Erklärungen von Fremdenfeindlichkeit in Deutschland [Social psychological analyses and explanations for hostility towards strangers in Germany]. *Zeitschrift für Sozialpsychologie, 32*, 59–79. doi: 10.1024//0044-3514.32.2.59

Walsh, T. M. (2013). *Implicit measures of homophobia and stigmatization of same-sex couples.* Unpublished master's thesis. Tempe, AZ.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95–116). New York, NY: Guilford Press.

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review, 19*, 1057–1064. doi: 10.3758/s13423-012-0295-x

Wuensch, K. L. (2012). Using SPSS to obtain a confidence interval for Cohen's *d*. Retrieved from http://core.ecu.edu/psyc/wuenschk/SPSS/CI-d-SPSS.pdf

Received October 12, 2015
Revision received February 22, 2017
Accepted February 27, 2017
Published online June 20, 2017

Sarah Teige-Mocigemba
Institut für Psychologie, Sozialpsychologie und Methodenlehre
Albert-Ludwigs-Universität Freiburg
79085 Freiburg
Germany
teige@psychologie.uni-freiburg.de

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

 © 2017 Hogrefe Publishing